DEPARTMENT OF ACTUARIAL STUDIES RESEARCH PAPER SERIES

The Predictive Distribution for a Poisson Claims Model

by

Glen Barnett and Angela Tong

Queries to gbarnett@efs.mq.edu.au

Research Paper No. 2008/01
December 2008

Division of Economic and Financial Studies
Macquarie University
Sydney NSW 2109 Australia

The Macquarie University *Actuarial Studies Research Papers* are written by members or affiliates of the Department of Actuarial Studies, Macquarie University. Although unrefereed, the papers are under the review and supervision of an editorial board.

Editorial Board:
Piet de Jong
Leonie Tickle

Copies of the research papers are available from the World Wide Web at:

http://www.actuary.mq.edu.au/research/research_papers

Views expressed in this paper are those of the author(s) and not necessarily those of the Department of Actuarial Studies.

The predictive distribution for a Poisson claims model

Glen Barnett and Angela Tong

Abstract

We give a derivation of the frequentist predictive distribution for a simple Poisson model, show how this simple model may be adapted to fit a claim count triangle, and demonstrate how the predictive distribution of aggregates of future claim counts may be computed for the adapted model. Quantiles and prediction intervals are readily obtained.

Quasi-Poisson models for incremental paid losses are commonly used in reserving and risk capital calculations. The results in this paper may readily be extended to that situation for a similar model to the one applied to claim counts.

Section 1: Introduction

This section describes two models – a very simple Poisson count with exposure model with a single parameter, which we extend to a more complex illustrative example, and secondly a model for a claim count triangle that depends on multiple Poisson parameters. The result for the predictive distribution for the Poisson may be applied to these situations, and quantiles of individual and aggregate forecasts may be obtained.

1.1 Poisson counts with known exposures

Consider the case where we have a set of Poisson counts with known exposures and a common Poisson rate per unit of exposure. That is, consider a vector of independent Poisson random variables $X_1, \ldots, X_n$, representing a set of counts with known exposures $k_1, \ldots, k_n$. $X_i \sim \text{Poisson}(\mu_i)$ where $\mu_i = k_i \cdot \lambda$. While λ is unknown, we can estimate it from the observations on the X's, $x_1, x_2, \ldots, x_n$.

We wish to obtain the predictive distribution of and additional observation, X_{n+1} (with known exposure k_{n+1}), conditional on the observed Poisson counts $(x_1, \ldots, x_n)$ and their exposures.

1.2 An illustrative example

Single-vehicle accidents occurring on non-holiday weekdays on two different stretches of a highway are recorded for several months. One stretch is long (length d_L), and the second is short (d_S). Data is observed for n time periods, of length (in number of non-holiday-weekdays) $t_1, t_2, \ldots, t_n$. Interest focuses on prediction of the number of single-vehicle accidents in the next time period (length t_{n+1}) for both stretches, and on the combined total across both stretches of highway.

Time:	t_1	t_2	…	t_n	t_{n+1}
Long stretch (d_L)	X_{L1}	X_{L2}	…	X_{Ln}	$X_{L,n+1}$
Short stretch (d_S)	X_{S1}	X_{S2}	…	X_{Sn}	$X_{S,n+1}$

Each count random variable, X_{ij}, is assumed to be independent Poisson(μ_{ij}) where $\mu_{ij} = \lambda \cdot d_i t_j$, $i = L,S$, j = 1, 2, …n+1. That is, the exposure is $k_{ij} = d_i t_j$ (measured in *km-days*), and the unknown event rate per unit of exposure is λ.

This illustrative example is a situation where techniques such as bootstrapping or Bayesian methods might previously have been used to obtain predictive distributions.

1.3 Payments-per-exposure and claims-per-exposure models

Many authors have looked at fitting a mean to incremental payments or incremental payments per exposure (generally adjusted for economic inflation) for each development. Names for methods based on this idea include Payments per Unit of Risk (Cumpston, 1976), Payments per Exposure (Taylor and Ashe, 1983), the Additive Method (Stanard, 1983), among others. Here we apply this approach to count triangles.

That is, we assume that the incremental number of claims (for example *claims reported* or *claims finalized*) in a given development for a particular accident period has a Poisson distribution whose mean depends only on the development period and the exposure for a given accident period. Given the known development periods and exposures, each count is assumed independent.

That is, for a triangle with s accident periods, the number of claims, X_{ij} in accident period i, development period j is assumed to be distributed as Poisson($k_i \cdot \mu_j$), for known exposure k_i.

1.3.1 The observed claims triangle

Table 1 below is a count triangle with exposures, a subset of data from Francis (2005), which is based on closed claim count data from the Texas Dept. of Insurance web site.

		DY		
AY	Exposure	0	1	2
1998	141.9	168	33	3
1999	141.4	117	42	6
2000	137.5	102	50	0
2001	176.7	185	0	0
2002	192.0	170	16	
2003	197.3	171		

Table 1: Count and exposure data

The observed counts beyond delay 2 were all zero, and we assume for the present illustration that development year 3 and beyond will remain zero for all years. Interest is then on prediction of the three unobserved future cells in the lower right of table 1. For insurers, interest will lie mainly in prediction of the row-sums of those future cells and prediction of the overall sum of the unobserved cells. In some circumstances, the sum of the counts by calendar (or payment) year may be useful.

Section 2: The predictive distribution for a Poisson random variable

The Bayesian predictive distribution for the Poisson with a conjugate (gamma) prior is a standard result, yielding a negative binomial predictive distribution.

A number of authors have considered frequentist prediction of Poisson observations, for example, Bain and Patel (1993) who give an indirect method of constructing prediction limits from tables for several different discrete distributions. In the case of the Poisson distribution, the prediction limits are obtained indirectly via tables of the binomial distribution. Patil (1960) describes finding negative binomial probabilities via binomial tables; indeed the predictive distribution result of this section might have been inferred from combining the information in Patil with the result in Bain and Patel (though less directly than the approach taken here). Datta et al. (2000), consider Bayesian prediction intervals with asymptotically matched coverage probabilities to frequentist prediction.

The bootstrap has become popular as a tool for generating prediction intervals. Harris (1989) compares bootstrap predictive distributions and fitted distributions, and shows that the bootstrap does better than fitted distributions in terms of Kullback-Leibler distance from the approximation to the predictive distribution he uses. In the case of the Poisson, Harris uses a Bayesian predictive distribution as the benchmark, (because the frequentist predictive distribution is not available).

Lawless and Fredette (2005) construct frequentist prediction intervals based on pivotal quantities, and include a Poisson example where approximate prediction intervals may be obtained from asymptotically pivotal quantities via simulation. They also compare with fitted distributions and show that their approach is better in terms of Kullback-Leibler distance.

2.1 The thinned Poisson Process

In this section we give a simple derivation of the predictive distribution for a Poisson random variable.

Consider a Poisson process at intensity rate λ, where there are two types of events. For each Poisson event, the event is either of type I (with probability π) or type II, (with probability $1-\pi$). The event-types are generated by a Bernoulli process, independent of the Poisson process. When interest focuses on one of the two event types, say the

type I events, this is sometimes called a *thinned Poisson process*. The selected events are a Poisson process with intensity $\pi\lambda$, independent of the Poisson process corresponding to the unselected events, which occurs at rate $(1-\pi)\lambda$. This is a standard result found in many basic texts (e.g. Ross, 2006, Proposition 5.2).

2.2 Poisson predictive distribution

Consider a Poisson process for a period of length $t_1 + t_2$ at intensity rate λ, and a thinning Bernoulli process with type-I probability $\pi = t_1/(t_1 + t_2)$. From the above thinned-Poisson results, conditional on λ, the number of events of the two types are independent Poisson random variables, with the number of type-I (selected) events $X_1|\lambda \sim \text{Poisson}(t_1 .\lambda)$, and the number of unselected events, $X_2|\lambda \sim \text{Poisson}(t_2 .\lambda)$.

Suppose we observe the number of selected events $X_1 = x_1$ but we do not know the number of unselected events. We want to find the distribution of X_2 given that we know the value of X_1. (Because we only have the information in X_1 to tell us about λ, this will be consistent with a range of values of λ, as described by the likelihood. Where we are unable to condition on λ, X_1 contains information on the mean of X_2.)

In the above thinned Poisson process, consider just the Bernoulli thinning process (the selection process). For a Bernoulli process, the number of events of type II until the x_1^{th} type I event is a standard result, so we immediately obtain that the distribution of $(X_2|X_1=x_1)$ is $\text{NegBin}[x_1, t_1/(t_1 + t_2)]$.

This establishes the frequentist predictive distribution for a Poisson random variable.

Note that some authors discuss the form of the negative binomial where the number of *trials*, rather than the number of *failures*, is the random variable. That negative binomial distribution is of the same basic form as the one discussed here, but shifted by the (known) number of successes. We use the "number of failures" form of the negative binomial throughout this paper.

2.3 The Poisson model

The result from the previous section can be extended to cases with multiple Poisson observations. From section 1.1, we have a vector of independent Poisson random variables $X_1, \ldots, X_n$, representing a set of counts with known exposures $k_1, \ldots, k_n$. $X_i \sim \text{Poisson}(\mu_i)$ where $\mu_i = k_i \cdot \lambda$. While λ is unknown, we can estimate it from the observations on the X's, $x_1, x_2, \ldots, x_n$.

Let $Y_n = X_1 + X_2 + \ldots + X_n$, $y_n = x_1 + x_2 + \ldots + x_n$, and $h_n = k_1 + k_2 + \ldots + k_n$.

The predictive distribution of X_{n+1} (with known exposure k_{n+1}), conditional on the observed Poisson counts $(x_1, \ldots, x_n)$, is given by $\text{NegBin}(y_n, h_n/h_{n+1})$.

2.3.1 Estimates and standard errors

First, we give the (well known) estimate of the parameter, λ, and its standard error.

$$\hat{\lambda} = y_n/h_n$$

$$\mathrm{Var}(\hat{\lambda}) = \mathrm{Var}(y_n/h_n) = (h_n)^{-2}\,\mathrm{Var}(y_n) = (h_n)^{-2}.h_n\,\lambda = \lambda/h_n\,.$$

This quantity is estimated by $\hat{\lambda}/h_n$, so the estimated standard error is $(\hat{\lambda}/h_n)^{1/2}$.

The estimate of the ith mean, μ_i is

$$\hat{\mu}_i = k_i \cdot \hat{\lambda}$$

with variance $k_i^2\,\mathrm{Var}(\hat{\lambda}) = k_i^2/h_n \cdot \lambda = (k_i/h_n)\,.\,\mu_i$ and so its estimated standard error is $(k_i/h_n \cdot \hat{\mu}_i)^{1/2}$.

2.3.2 Prediction of a future count

The usual prediction of X_{n+1} is given by
$\hat{X}_{n+1} = \hat{\mu}_{n+1} = k_{n+1} \cdot \hat{\lambda} = k_{n+1} \cdot y_n / h_n$.

We could also derive the predictive mean $\hat{\mu}_{n+1}$ by considering the predictive distribution of $(X_{n+1} \mid Y_n = y_n, h_n, k_{n+1})$, which is negative binomial. The mean and variance of a NegBin[x, p] are $x(1-p)/p$ and $x(1-p)/p^2$ respectively (so for a negative binomial, the variance is obtained by dividing the mean by p.)

When predicting X_{n+1}, p is h_n/h_{n+1}.

Thus, $\mathrm{E}(\hat{X}_{n+1}|y_n) = y_n\,(k_{n+1}/h_n) = k_{n+1} \cdot y_n/h_n = \hat{\mu}_{n+1}$, as expected.

The variance of $\hat{\mu}_{n+1}$,

$$\mathrm{Var}(\hat{\mu}_{n+1}) = \mu_{n+1} \cdot k_{n+1}/h_n\,,$$

which is estimated by $\hat{\mu}_{n+1} \cdot k_{n+1}/h_n$.

The predictive variance of $(X_{n+1}|\,y_n)$ is thus:

$$\begin{aligned}
\mathrm{Var}(X_{n+1}|\,y_n) &= \hat{\mu}_{n+1}\,(h_{n+1}/h_n) \qquad\qquad \text{(again, variance = mean/}p\text{)}\\
&= \hat{\mu}_{n+1}\,(1 + k_{n+1}/h_n)\\
&= \hat{\mu}_{n+1} + \hat{\mu}_{n+1}\;k_{n+1}/h_n\\
&= \text{process variance} + \text{parameter uncertainty,}
\end{aligned}$$

again, as expected.

2.4 Illustrative Example

We now return to the illustrative example of section 1.2, which is a situation where the bootstrap, or possibly Bayesian methods would have been employed to obtain prediction intervals.

In section 1.2, each count random variable, X_{ij}, was assumed to be independent Poisson(μ_{ij}) where $\mu_{ij} = \lambda \cdot d_i t_j$, $i = L,S$, j = 1, 2, …n+1. That is, the exposure for X_{ij} is $k_{ij} = d_i t_j$, and the unknown event rate per unit of exposure is λ.

Let $D = \sum_{i=L}^{S} d_i$, $T_n = \sum_{j=1}^{n} t_j$, $X_{\bullet,n+1} = X_{L,n+1} + X_{S,n+1}$, and $k_{\bullet,n+1} = k_{L,n+1} + k_{S,n+1} = \mathrm{D} \cdot \mathrm{t}_{n+1}$.

Observed data, ($x_{L1}, x_{S1}, x_{L2}, x_{S2}, \ldots, x_{Ln}, x_{Sn}$) are obtained.
Let $Y_n = \sum_{i=L}^{S}\sum_{j=1}^{n} X_{ij}$, $y_n = \sum_{i=L}^{S}\sum_{j=1}^{n} x_{ij}$ and $h_n = \sum_{i=L}^{S}\sum_{j=1}^{n} k_{ij} = \sum_{i=L}^{S} d_i \sum_{j=1}^{n} t_j = D \cdot T_n$

Then in the same fashion as the previous results, we have

$(X_{i,n+1} | y_n, h_n, k_{i,n+1}) \sim \mathrm{NB}[y_n, h_n/(h_n + k_{i,n+1})]$, for $i = L,S$

and similarly

$(X_{\bullet,n+1} | y_n, h_n, k_{\bullet,n+1}) \sim \mathrm{NB}[y_n, h_n/h_{n+1}]$ for the next period aggregate.

Consequently, we can give quantiles and one- or two-sided prediction intervals of any desired coverage probability (up to the discreteness of the negative binomial).

Note that because they share the same estimate of λ, $\hat{\lambda} = y_n/h_n$, these predictive distributions will not be independent.

Section 3: A Claims-per-exposure model

We now proceed to an example for which the aggregate predictive distribution does not seem to be directly computable analytically. Analytical results for aggregate predictive distributions are not available, but these may be computed in other ways (sections 3.3.1 and 3.3.2 discuss two approaches).

3.1 Predictive distributions for the claims triangle with exposures

Refer again to Table 1 (section 1.3.1), which is a count triangle with exposures.
As in section 1.3, for a triangle with s accident periods, the number of claims, X_{ij} in accident period i, development period j is assumed to be distributed as Poisson($k_i \cdot \mu_j$), for known exposure k_i. We assume that development year 3 and beyond will remain zero for all years.

The model for each column (development year) is, in effect, the Poisson exposure model described in section 1.1 and for which the predictive distribution is given in section 2.3.

3.1.1 Poisson parameter estimates

	DY0	DY1	DY2
$y = \sum x$	913	141	9
$h = \sum k$	986.8	789.5	597.5
$\hat{\lambda} = y/h$	0.925213	0.178594	0.015063

Table 2: Poisson parameter estimates

The parameter estimate for DY0, though given in Table 2 above, is not required for prediction with this model; however, it can be used to compute fitted values and residuals.

3.1.2 Calculation of parameters of predictive distributions

DY1:

$(X_{2003,1}|y_1 = 141) \sim \text{NegBin}(141, 789.5/986.8)$ $\quad (p \cong 0.800)$

Hence mean = $141 \times 0.2/0.8 = 35.2366$, and std.dev. = $\sqrt{(35.2366/0.8)} = 6.64$

DY2:

$(X_{2002,2}|y_2 = 9) \sim \text{NegBin}(9, 597.5/789.5)$ $\quad (p \cong 0.757)$

$(X_{2003,2}|y_2 = 9) \sim \text{NegBin}(9, 597.5/(597.5+197.3))$ $\quad (p \cong 0.752)$

Table 3 below contains observations and forecasts. Below the observed values are fitted values, and below the forecasts are standard deviations.

		DY			
AY	Exposure	0	1	2	
1998	141.9	168	33	3	
		131.3	25.3	2.1	
1999	141.4	117	42	6	
		130.8	25.3	2.1	
2000	137.5	102	50	0	
		127.2	24.6	2.1	
2001	176.7	185	0	0	**AY total**
		163.5	31.6	2.7	Pr. St.Er
2002	192.0	170	16	2.89	2.89
		177.6	34.3	1.95	1.95
2003	197.3	171	35.24	2.97	38.21
		182.5	6.64	1.99	6.93
	DY total outstanding		35.24	5.86	41.10
	Pred Std Error		6.64	3.11	7.33
			2004	2005	
	CY total outstanding		38.13	2.97	
	Pred Std Error		6.92	1.99	

Table 3: Predictive means and standard deviations for each cell and period totals.

All of the numbers in the table can be computed without reference to the negative binomial, just by using the Poisson model and basic statistical results. However, that will not suffice if we wish to know a quantile (such as a 95th percentile), or to give a prediction interval for one of the forecasts; the negative binomial distribution can be used in that circumstance.

In figure 1 below is the predictive probability function for one of the future values.

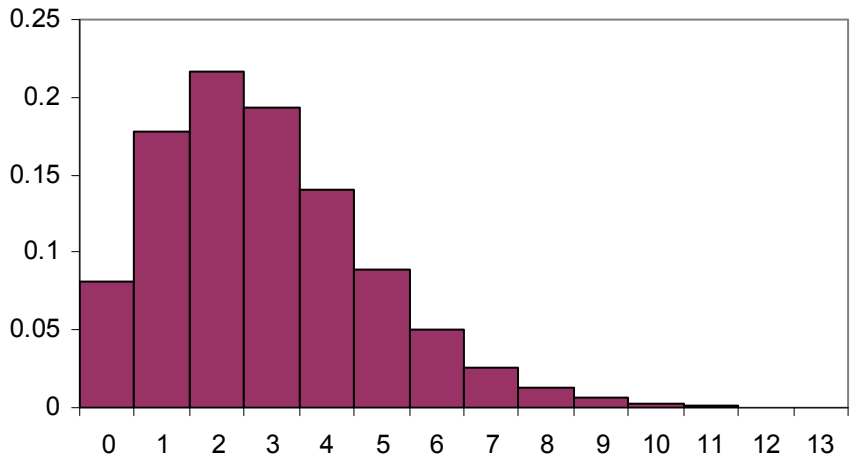

Figure 1: Negative binomial predictive distribution probability function of the count in accident year 2002, development year 2.

3.2 The distribution of column totals under the model

Let $y_j = x_{1,j} + x_{2,j} \ldots + x_{s-j,j}$ be the observed number of claims in *development j* . Let k_1, k_2, …, k_{s-j} be the corresponding exposures, and let $h_j = k_1 + k_2 \ldots + k_{s-j}$.

3.2.1 Parameter estimates and standard error

This constant-count-per-exposure Poisson model for each development can be fitted using a GLM. However, the model is simple enough that we can write down the parameter estimates directly.

$$\hat{\mu}_j = y_j/h_j$$

The standard error for $\hat{\mu}_j$ can also be easily computed:

$$\mathrm{var}(\hat{\mu}_j) = \mathrm{var}(y_j)/h_j^2 = h_j.\mu_j/h_j^2 = \mu_j/h_j\,.$$

Thus the standard error for $\hat{\mu}_j$ is estimated by $(\hat{\mu}_j/h_j)^{1/2}$.

3.2.2 Predictive distribution

For a particular development period *j*, we observe (*s–j*) Poisson random variables, X_{1j}, X_{2j}, …, $X_{s-j,j}$ with corresponding exposures k_1, k_2, …,k_{s-j}, which are known. We also

know the exposure k_t for the Poisson random variable $X_{t,j}$, which represents the incremental number of claims in a future cell, $t > s–j$.

Thus the predictive distribution for X_{tj} is given by

$$(X_{t,j}|\, y_j) \sim \text{NegBin}[y_j,\, h_j/(h_j+k_t)].$$

Further, given μ_j, the total future claims in a given development is Poisson($K_j \cdot \mu_j$), where $K_j = k_{s-j+1} + k_{s-j+2} + \ldots + k_s$ represents the sum of accident-period exposures for those accident periods corresponding to future payment periods in this development period. Note that $h_j + K_j = h_s$.

Consequently, the predictive distribution of the total outstanding claim count, $T_j = X_{s+1-j,j} + \ldots + X_{s,j}$ within a given development is

$$(T_j|\, y_j) \sim \text{NegBin}[\, y_j,\, h_j/h_s\,].$$

The individual prediction errors within a development period are not independent, since they share parameter estimates. However, across developments, the prediction errors are independent.

3.3 Aggregate future claim count distributions

As we have seen, we can compute the predictive distribution of the future claim count for accident period i, development period j, and also the predictive distribution of the aggregate of future claims across all accident periods in development period j.

However, we do not have simple expressions for the predictive distribution for *all* unreported claims (though its mean, variance and some other characteristics are readily computed), nor for individual accident periods. The reason for this is that under the model, different parameters apply for different development periods. Our derivation of the negative binomial for the predictive distribution was based on Poisson random variables which shared a common rate per unit of exposure. With two or more unrelated event rates, the required distribution of the aggregate is a convolution of independent negative binomials with different p parameters.

While we cannot give simple closed-form analytical expressions for the aggregate predictive distribution of future claim counts for this multi-parameter model, we can obtain the required distribution to a desired level of accuracy relatively easily via simulation or numerical convolution of the individual negative binomial distributions.

The predictive distribution for each development period is independent of the other developments. This allows us to rapidly simulate the predictive distribution of future claims by simulating from the predictive distribution of aggregate claims for each development and adding the simulated values to obtain the overall aggregate. Similarly we can simulate the independent components for an individual accident period to obtain simulated values from the aggregate for that accident year, or similarly for a given calendar year.

It is also possible to compute aggregate distributions via numerical convolution. However, such numerical convolution should be carefully implemented to be practical as the mean of each variable and the number of variables in the convolution grows large; there are a variety of algorithms available for convolution of negative binomial random variables, for example via the Fast Fourier Transform, a minor modification of the algorithm of O'Cinneide and Schneider (1987), or the approach of Furman (2007), all of which will be faster in general than taking a naïve direct sum-of-products-of-probabilities calculation as each new variable is added in. For small problems, however, brute-force convolution methods may be quite practical.

3.3.1 Calculations via simulation

The development year total outstanding count was simulated for development years 1 and 2 and pairs of simulated values added together to obtained simulated values from the distribution of the aggregate under the model. This was performed in the statistical package *R* (via the `rnbinom` command) – 10000 random values were generated from the predictive distribution for each development year total and added to give 10000 values for the total future count.

Nominal Percentile Rank	Percentile (x)	Sample cdf F(x)
50%	41	52.86%
75%	46	76.82%
95%	54	95.63%
99.5%	62	99.59%

Mean	41.2
Sd	7.35

Table 4: Percentiles from simulated distribution of aggregate

Table 4 above gives a set of sample quantiles from the simulated distribution of the outstanding aggregate count. Note that the simulated sample mean and standard deviation are close to their theoretical values. Figure 2 below is a histogram of the simulated values from the aggregate outstanding count.

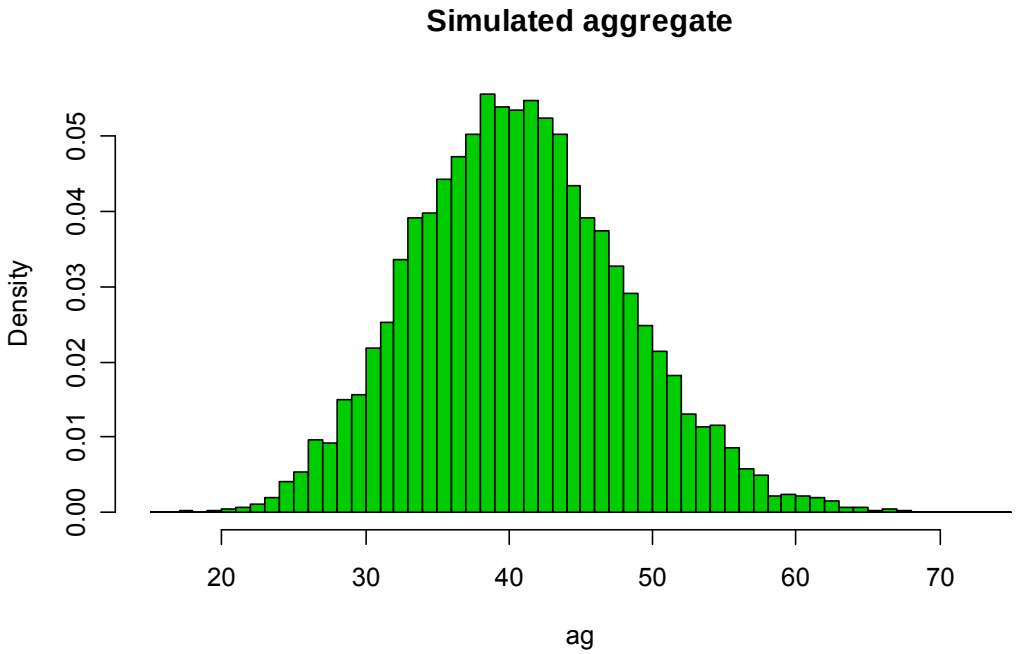

Figure 2: Histogram of simulated values for the aggregate outstanding count

3.3.2 Calculations via numerical convolution

As mentioned earlier, we can also compute the convolutions numerically. In this case, since it was such a small problem, the convolution was done by naïve direct sum-of-products, involving truncation of the smaller negative binomial (mean 5.86) at 30 and above (for which the probabilities were below 5×10^{-7}).

Percentile calculations from numerical convolution for the aggregate outstanding count are given in Table 5 below, again, along with the computed mean and standard deviation for comparison with the theoretical values. (The calculations performed were more precise than required for the degree of accuracy displayed, and so match the theoretical values.)

Nominal Percentile Rank	Percentile (x)	Actual cdf F(x)
50%	41	53.70%
75%	46	77.47%
95%	54	96.03%
99.5%	61	99.50%

mean	41.1
sd	7.33

Table 5: Percentiles from numerical convolution for distribution of the aggregate

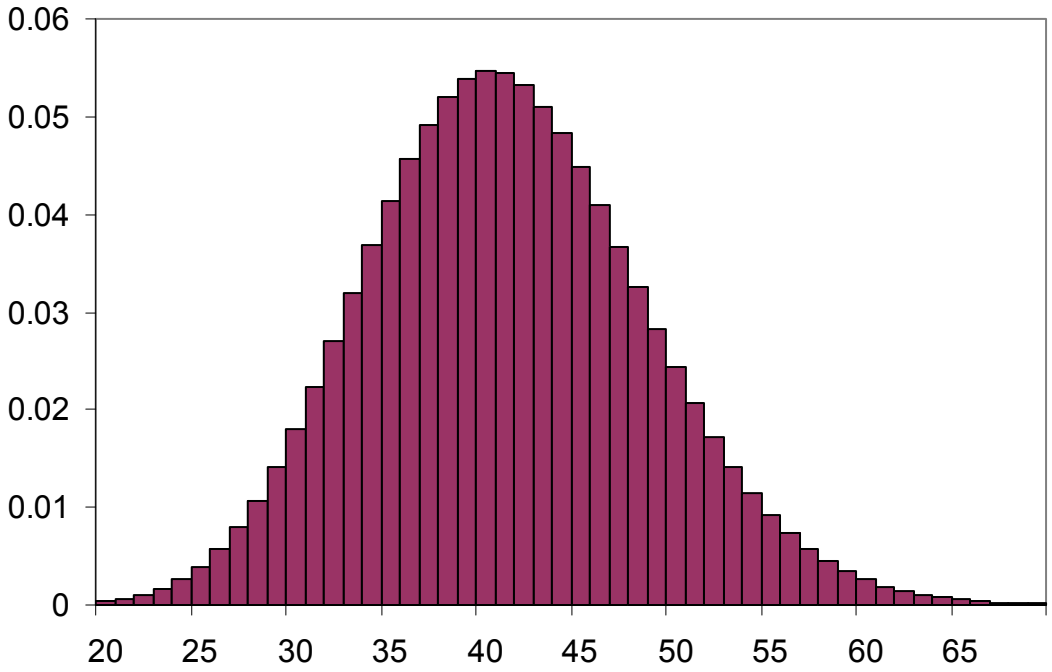

Figure 3: Probability function for the predictive distribution of total outstanding count

Figure 3 above shows the predictive distribution from numerical convolution for the aggregate outstanding count. It again is very similar to (but smoother than) the one obtained from simulation.

3.4 The quasi-Poisson model

Although our discussion is based on a Poisson model, the results apply to the quasi-Poisson (overdispersed Poisson) model with an appropriate change of scale. This is established by uniqueness properties in the exponential family (e.g. Venter, 2007), or as Schmidt (2002) shows via the MGF, under a simple condition, the distribution of a regular exponential dispersion family overdispersed Poisson is a scaled Poisson distribution. That is, the overdispersed-Poisson exponential family model may be regarded as simply a scaled Poisson model.

Consequently, if the quasi-Poisson model is used for claim payments, it is equivalent to assuming a constant severity and a Poisson (or possibly scaled-Poisson) count.

This means that the derivations here apply equally to paid losses that are modeled with a quasi-Poisson GLM, after appropriate scaling.

3.5 Diagnostics

As with any model, it is necessary to check model assumptions, and to focus most closely on those assumptions most likely to fail.

3.5.1 The Poisson variance assumption

One of the most critical assumption for these calculations is probably the Poisson "*variance = mean*" assumption. If we fit the GLM (here via *R*), we obtain (in part) the following output:

```
Residual deviance:  141.43  on 12  degrees of freedom
```

This is well beyond the bounds of what we might reasonably see with Poisson data.

Fortunately, taking an overdispersed-Poisson assumption merely involves a scaling (as described earlier). We estimate the variance parameter, $\hat{\phi}$= 141.43/12 = 11.7858.

Hence the scaling factor is 3.433. The resulting scaled Poisson has a scaled negative binomial predictive distribution.

3.5.2 Mean specification

The most obvious ways for the model to fail in its description of the mean are for the exposures to fail to capture the accident year trend changes, and for there to be calendar year (payment year) trends – in practice calendar period trend changes turn out to be surprisingly common in claim count data.

Below is a plot of Pearson standardized residuals vs calendar year. As mentioned earlier, CY trends do occur with count data, and it appears to be the case here. There's some distinct patterns in the calendar year direction – 2001 is all above zero, while 2002 and 2003 are all below it (though there's only a few observations in each of these years). This suggests the present model may be inadequate.

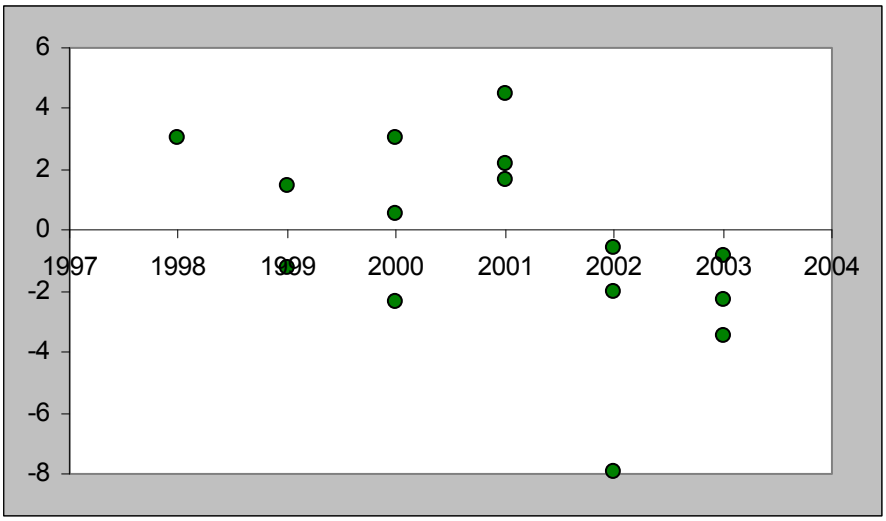

Figure 4: Deviance residuals vs calendar (payment) year

We should also examine the plot of residuals vs accident year, which if there are no calendar period effects would tell us about the suitability of the exposure measure. Figure 5 below shows something of a downward trend. Note however that for this

data, because there are only 3 different development years of data, the accident year and calendar year effects are fairly strongly confounded, so it is difficult to separate the effects in the two directions, at least with this model. All we can say without a better model is that there are some indications of model inadequacy in at least one of the accident year and calendar year directions.

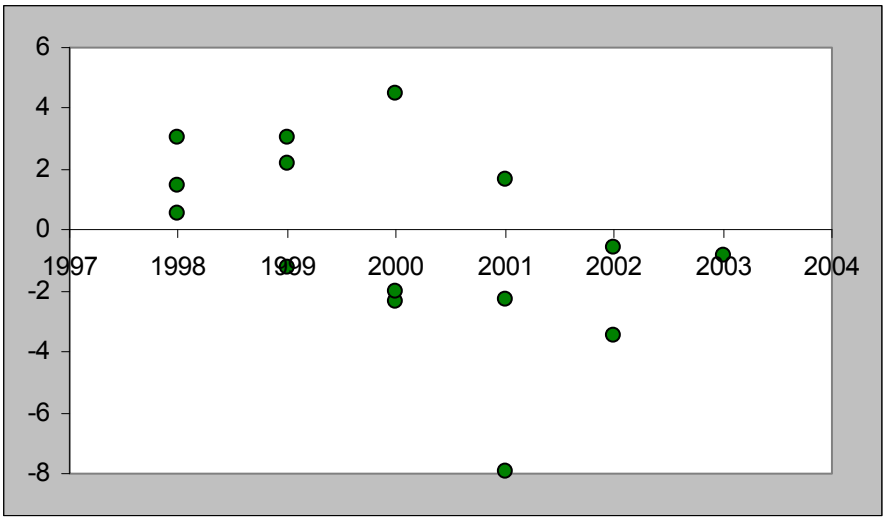

Figure 5: Deviance residuals vs accident year

A plot against the development year can show no lack of fit in the mean, since the model is fully parameterized in that direction. It is possible to use such a plot to assess whether the specification of the relative variances is reasonable – if variance is not proportional to mean, it shows up as changing spread in the plot against development year. However, aside from the single low outlier, corresponding to a zero count in accident year 2001, development year 1, there is nothing of consequence in that plot and it is not included here.

3.6 The use of the count-per-exposure model

Note that this form of Poisson model (even in the overdispersed form) is not necessarily a particularly good model of claim counts for a number of reasons. Firstly, claim counts often show calendar period effects (for example due to social, economic, legislative, judicial or intra-company factors influencing either probability or the timing of some claims). Secondly, the model is somewhat overparameterized, since it makes no attempt to relate the levels across developments, even though the trends across development are often quite smooth. Note that the popular chain ladder and quasi-Poisson two-way main effects models share these drawbacks, and have an even worse problem with overparameterization.

However, this model does serve a number of useful purposes. Firstly, in some circumstances it is at least a reasonable model. Secondly, it may be a useful starting point for modeling count triangles. Thirdly, because we can compute the predictive distribution of future claims in each development, we can use this as a tool to evaluate the suitability of proposed approaches for calculating predictive distributions, such as bootstrapping, since deficiencies that we might reveal in this relatively simple case will carry over to related but more complex models.

References:

Bain, L.J. and Patel, J.K. (1993) Prediction intervals based on partial observations for some discrete distributions, *IEEE Transactions on Reliability,* 42, No. 3, 459-463.

Cumpston, J.R., 1976, Payments per unit of risk claims models, *General Insurance Bulletin* 1, 8-11.

Datta, G.S., Mukerjee, R., Ghosh, M. & Sweeting, T.J. (2000), Bayesian prediction with approximate frequentist validity, *Annals of Statistics*, 28, No. 5 1414-1426

Francis, L. (2005) Introduction to Generalized Linear Models, *CAS Predictive Modeling Seminar* (www.casact.org/education/specsem/f2005/handouts/GLM1.ppt)

Furman, E. (2007) On the convolution of the negative binomial random variables, *Statistics and Probability Letters,* 77, 169-172

Harris, I. R. (1989). Predictive fit for natural exponential families. *Biometrika* **76** 675–684.

Lawless, J.F. and Fredette, M. (2005) Frequentist prediction intervals and predictive distributions, *Biometrika*, 92, No. 3, 529-542

O'Cinneide and Schneider (1987) An algorithm for computing the convolution of Poisson, negative binomial and geometric distributions, *Operations Research*, 35, No. 3, 445-449.

Patil, G. P. (1960) On the Evaluation of the Negative Binomial Distribution with Examples, *Technometrics*, 2, No. 4, 501–505.

Ross, S. M. (2006) *Introduction to Probability Models*, 9th Ed, Academic Press, Orlando, FL.

Schmidt, K. (2002), A note on the overdispersed Poisson family, *Insurance: Mathematics and Economics*, 30, p21-25

Stanard, J.N. (1983), A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniqes, *Proceedings of the Casualty Actuarial Society LXXII*

Taylor, G.C. and Ashe, F.R. (1983) Second Moments of Estimates of Outstanding Claims. *Journal of Econometrics* 23, 37-61.

Venter, G. (2007) Generalized Linear Models Beyond the Exponential Family with Loss Reserve Applications, *CAS E-Forum, Summer 2007*